

Computing likelihoods in the stochastic integrate-and-fire model: numerical methods

Jonathan W. Pillow
Gatsby Computational Neuroscience Unit, UCL
17 Queen Square, London WC1N 3AR, UK
pillow@gatsby.ucl.ac.uk

Recent work has examined the estimation of models of stimulus-driven neural activity in which a linear filtering process is followed by a nonlinear, probabilistic spiking stage. We analyze the estimation of one such model for which this nonlinear step is implemented by a noisy, leaky, integrate-and-fire mechanism with a spike-dependent after-current. We have formulated this problem in terms of maximum likelihood estimation: a full discussion of the problem is contained in [1, 3]). Here we present detailed numerical methods related to computing the likelihood function using the Fokker-Planck equation, excerpted from [2]. This model was first applied to neuronal data in [4].

1 Defining the model and likelihood function

We consider a model for which the (dimensionless) subthreshold voltage variable V evolves according to

$$dV = \left(-gV(t) + \vec{k} \cdot \vec{x}(t) + \sum_{j=0}^{i-1} h(t - t_j) \right) dt + \sigma N_t, \quad (1)$$

and resets to V_r whenever $V = 1$. Here, g denotes the leak conductance, $\vec{k} \cdot \vec{x}(t)$ the projection of the input signal $\vec{x}(t)$ onto the linear kernel \vec{k} , h is an “afterpotential,” a current waveform of fixed amplitude and shape whose value depends only on the time since the last spike t_{i-1} , and N_t is an unobserved (hidden) noise process with scale parameter σ . Without loss of generality, the “leak” and “threshold” potential are set at 0 and 1, respectively, so the cell spikes whenever $V = 1$, and V decays back to 0 with time constant $1/g$ in the absence of input. The dynamical properties of this type of “spike response model” have been extensively

studied [5]; for example, it is known that this class of models can effectively capture much of the behavior of apparently more biophysically realistic models (e.g. Hodgkin-Huxley).

Our problem now is to estimate the model parameters $\{\vec{k}, \sigma, g, V_r, h\}$ from a sufficiently rich, dynamic input sequence $\vec{x}(t)$ together with spike times $\{t_i\}$. A natural choice is the maximum likelihood estimator (MLE), which is easily proven to be consistent and statistically efficient here. To compute the MLE, we need to compute the likelihood and develop an algorithm for maximizing it.

In the noiseless case [6], the voltage trace during an interspike interval $t \in [t_{i-1}, t_i]$ is given by the solution to equation (1) with $\sigma = 0$:

$$V_0(t) = V_r e^{-gt} + \int_{t_{i-1}}^t \left(\vec{k} \cdot \vec{x}(s) + \sum_{j=0}^{i-1} h(s - t_j) \right) e^{-g(t-s)} ds, \quad (2)$$

which is simply a linear convolution of the input current with a negative exponential. It is easy to see that adding Gaussian noise to the voltage during each time step induces a Gaussian density over $V(t)$, since linear dynamics preserve Gaussianity [7]. This density is uniquely characterized by its first two moments; the mean is given by (2), and its covariance is $\sigma^2 E_g E_g^T$, where E_g is the convolution operator corresponding to e^{-gt} . Note that this density is highly correlated for nearby points in time, since noise is integrated by the linear dynamics. Intuitively, smaller leak conductance g leads to stronger correlation in $V(t)$ at nearby time points. We denote this Gaussian density $G(\vec{x}_i, \vec{k}, \sigma, g, V_r, h)$, where index i indicates the i th spike and the corresponding stimulus chunk \vec{x}_i (i.e. the stimuli that influence $V(t)$ during the i th interspike interval).

Now, on any interspike interval $t \in [t_{i-1}, t_i]$, the only information we have is that $V(t)$ is less than threshold for all times before t_i , and exceeds threshold during the time bin containing t_i . This translates to a set of linear constraints on $V(t)$, expressed in terms of the set

$$C_i = \bigcap_{t_{i-1} \leq t < t_i} \left\{ V(t) < 1 \right\} \cap \left\{ V(t_i) \geq 1 \right\}.$$

Therefore, the likelihood that the neuron first spikes at time t_i , given a spike at time t_{i-1} , is the probability of the event $V(t) \in C_i$, which is given by

$$L_{\vec{x}_i, t_i}(\vec{k}, \sigma, g, V_r, h) = \int_{C_i} G(\vec{x}_i, \vec{k}, \sigma, g, V_r, h),$$

the integral of the Gaussian density $G(\vec{x}_i, \vec{k}, \sigma, g, V_r, h)$ over the set C_i .

Spiking resets V to V_r , meaning that the noise contribution to V in different interspike intervals is independent. This ‘‘renewal’’ property, in turn, implies that the density over $V(t)$ for an entire experiment factorizes into a product of conditionally independent terms, where each of these terms is one of the Gaussian integrals derived above for a single interspike

interval. The likelihood for the entire spike train is therefore the product of these terms over all observed spikes. Putting all the pieces together, then, the full likelihood is

$$L_{\{\vec{x}_i, t_i\}}(\vec{k}, \sigma, g, V_r, h) = \prod_i \int_{C_i} G(\vec{x}_i, \vec{k}, \sigma, g, V_r, h), \quad (3)$$

where the product, again, is over all observed spike times $\{t_i\}$ and corresponding stimulus chunks $\{\vec{x}_i\}$.

Now how do we actually compute the likelihood? This is a nontrivial problem: we need to be able to quickly compute integrals of multivariate Gaussian densities G over simple but high-dimensional orthants C_i . The technique we employ can be termed “density evolution” [8, 9]. The method is based on the following well-known fact from the theory of stochastic differential equations [7]: given the data (\vec{x}_i, t_{i-1}) , the probability density of the voltage process $V(t)$ up to the next spike t_i satisfies the following partial differential (Fokker-Planck) equation:

$$\frac{\partial P(V, t)}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial V^2} + g \frac{\partial [(V - V_{eq}(t))P]}{\partial V}, \quad (4)$$

under the boundary conditions

$$\begin{aligned} P(V, t_{i-1}) &= \delta(V - V_r), \\ P(V_{th}, t) &= 0; \end{aligned} \quad (5)$$

where $V_{eq}(t)$ is the instantaneous equilibrium potential:

$$V_{eq}(t) = \frac{1}{g} \left(\vec{k} \cdot \vec{x}(t) + \sum_{j=0}^{i-1} h(t - t_j) \right). \quad (6)$$

Moreover, the conditional firing rate $f(t)$ satisfies

$$\int_{t_{i-1}}^t f(s) ds = 1 - \int P(V, t) dV. \quad (7)$$

Thus standard techniques for solving the drift-diffusion evolution equation (4) lead to a fast method for computing $f(t)$. Finally, the likelihood $L_{\{\vec{x}_i, t_i\}}(\vec{k}, \sigma, g, V_r, h)$ is simply $f(t_i)$.

2 Numerical Methods for Fokker-Planck Equation

In order to compute the likelihood function $L_{\{\vec{x}_i, t_i\}}(\vec{k}, \sigma, g, V_r, h)$, we used a second-order numerical method for solving the Fokker-Planck (FP) equation (eq. 4). This equation describes the time evolution of $P(V, t)$, the probability density over sub-threshold voltage V at time t , as a function of the input and model parameters.

Our general approach involves discretizing V so that we can represent $P(V, t^*)$ at a fixed time t^* by a set of discrete values. We then propagate this density forward in time using the FP equation to obtain $P(V, t^* + \Delta t)$, the probability over V at the next time step. Intuitively, the likelihood of a spike occurring during the interval $[t^*, t^* + \Delta t]$ is given by the amount of probability mass which leaks over the (absorbing) boundary at threshold ($V = 1$) during this time step.

We now describe the density propagation algorithm in detail. Let $\{v_i\}_{i=1}^n$ denote the discretization over V , consisting of n evenly spaced bins with a separation of Δv . We let $v_n = 1$ (threshold) and set v_1 to some voltage sufficiently low that we can represent $P(V)$ accurately at all time points. We will use i to index voltage and j to index time, so p_i^j denotes the probability mass associated with the i th bin of the voltage discretization and time bin j . And, in a slight abuse of notation, we will use p^j to refer to the entire density over voltage at the j th time step.

We initialize the algorithm with a density p^1 , computed a short time after the most recent spike, when the subthreshold probability density over V is still well-approximated by a Gaussian. This initial density is given by

$$p_i^1 = N(V_0(t_1), \frac{1}{2g}(1 - e^{-2gt_1})\sigma^2), \quad (8)$$

the standard Gaussian density with mean $V_0(t_1)$ and variance $\frac{1}{2g}(1 - e^{-2g\tau})\sigma^2$, defined on the grid points v_i . Note that $V_0(t_1)$ is the noiseless voltage at time t_1 since the most recent spike (eq. 2).

Recall that the FP equation (eq. 4) for the model is given by

$$\begin{aligned} \frac{\partial P(V, t)}{\partial t} &= \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial V^2} + g \frac{\partial[(V - V_{eq}(t))P]}{\partial V}, \\ &= \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial V^2} + g(V - V_{eq}(t)) \frac{\partial P}{\partial V} + gP, \end{aligned} \quad (9)$$

where $V_{eq}(t)$ is the instantaneous reversal potential at time t (eq 6). We solve this equation using a scheme related to the Crank-Nicolson method for solving diffusive PDEs [10]¹. This involves substituting discrete approximations for the partial derivatives as follows:

$$\begin{aligned} \frac{p_i^{j+1} - p_i^j}{\Delta t} &= \frac{\sigma^2}{2} \left[\frac{(p_{i+1}^{j+1} - 2p_i^{j+1} + p_{i-1}^{j+1}) + (p_{i+1}^j - 2p_i^j + p_{i-1}^j)}{2(\Delta v)^2} \right] \\ &\quad + g(v_i - V_{eq}(t)) \left[\frac{(p_{i+1}^{j+1} - p_{i-1}^{j+1}) + (p_{i+1}^j - p_{i-1}^j)}{4(\Delta v)} \right] + g \frac{p_i^{j+1} + p_i^j}{2}. \end{aligned} \quad (10)$$

Note that the right-hand-side derivatives are evaluated by averaging over partial derivatives at the j th and $j + 1$ st time steps, leading to a method which is second-order accurate in V and t .

¹Note that the Crank-Nicolson method is unconditionally stable for purely diffusive PDEs, although not so if a drift term is included. Instabilities may therefore arise if Δt is too large relative to Δv .

For the sake of clarity, we can rewrite (10) as a sparse matrix equation, which can be solved efficiently in $o(n)$ operations. We have:

$$\frac{1}{\Delta t}(p^{j+1} - p^j) = \frac{\sigma^2}{4(\Delta v)^2}D''(p^{j+1} + p^j) + \frac{g}{4(\Delta v)}D'(V - V_{eq}(t))(p^{j+1} + p^j), \quad (11)$$

where D' and D'' are tri-diagonal matrices corresponding to derivative and second-derivative operators (with the values $[-1 \ 0 \ 1]$ and $[1 \ -2 \ 1]$ along the main diagonals, respectively), V is a diagonal matrix with the grid points $\{v_i\}$ along the diagonal, and $V_{eq}(t_j)$ is a scalar that depends on the input during the current time step. By collecting like terms, this equation can be simplified to

$$\left(A_1 - BV_{eq}(t_j)\right)p^{j+1} = \left(A_2 + BV_{eq}(t_j)\right)p^j, \quad (12)$$

where A_1 , A_2 and B are tri-diagonal matrices that can be computed in advance:

$$\begin{aligned} A_1 &= I - \frac{\sigma^2 \Delta t}{4(\Delta v)^2}D'' - \frac{g \Delta t}{4(\Delta v)}D'V \\ A_2 &= I + \frac{\sigma^2 \Delta t}{4(\Delta v)^2}D'' + \frac{g \Delta t}{4(\Delta v)}D'V \\ B &= \frac{g \Delta t}{4(\Delta v)}D', \end{aligned} \quad (13)$$

where I is the $n \times n$ identity matrix. We used a special routine written in C to solve this equation for p^{j+1} on each time step, which effects the density propagation.

Of course, we must also specify the correct boundary conditions (eq. 5) to ensure that probability mass leaks only one way across the spike threshold, which we use to compute $p(\text{spike})$ during each time step. We enforce the upper (absorbing) boundary condition by replacing the n th columns of the D' and D'' matrices with the n th column of the identity matrix (i.e. zero except for 1 in the n th position), which conserves probability mass in the last bin and prevents drift or diffusion from p_n to p_{n-1} . We enforce the lower (reflecting) boundary condition by adding to the first entry of D' and D'' so that first column sums to 1, which ensures that probability mass is conserved at the lower boundary (i.e. it doesn't leak out of the range of $\{v_i\}$).

After having initialized the density at p^1 , we perform density propagation (computing $V_{eq}(t)$ at each time step and solving equation 12) until we reach the next spike time t_k . Here, p_n^k gives the cumulative probability of a spike having occurred by time t_k , and the likelihood of a spike occurring at t_k is $\frac{p_n^k - p_n^{k-1}}{\Delta t}$.

3 The Gaussian process $V(t)$

Here we derive discrete and continuous solutions for the mean and variance of $V(t)$, the membrane potential of the IF model, in the absence of spiking. $V(t)$ is a Gaussian (Ornstein-Uhlenbeck) process, and therefore completely characterized by its mean and variance.

Mean

The mean, $\mu(t)$, of the Gaussian (governing the evolution of $P(V)$, the density over membrane potential) is equal to the solution of the noiseless version of the integrate-and-fire dynamics, on the interval $[0, t_i]$:

$$\frac{dV}{dt} = -gV(t) + \vec{k} \cdot \vec{x}(t), \quad (14)$$

with initial data

$$V(0) = V_r.$$

Thus,

$$\mu(t) = V_r e^{-gt} + \int_0^t (\vec{k} \cdot \vec{x}(s)) e^{-gs} ds.$$

To simplify notation, we can rewrite $\mu(t)$ in operator form:

$$\mu(t) = E_g [V_r \delta(0) + k \cdot \vec{x}(t)], \quad (15)$$

where E_g is the convolution operator corresponding to e^{-gt} .

If we consider the problem discretized in time bins of width Δt and set $\alpha = e^{-g\Delta t}$, the operator E_g can be written as a matrix:

$$E_g = \begin{bmatrix} 1 & & & & & \\ \alpha & 1 & & & & \\ \alpha^2 & \alpha & 1 & & & \\ \vdots & & & \ddots & & \\ \alpha^n & \dots & \alpha^2 & \alpha & 1 & \end{bmatrix}. \quad (16)$$

The first row corresponds to the filtering during the first time bin and the n th row corresponds to filtering for the n th time bin, or $t = n(\Delta t)$ of the solution. Analytically, for continuous time, we can express the mean as:

$$\mu(t) = \frac{1}{g}(1 - e^{-gt})I,$$

if I is a (constant) injected current.

Covariance

The covariance matrix Λ for $V(t)$ is given by the outer product of E_g with itself (this is true for any linear operator applied to a Gaussian random variable): $\Lambda = E_g E_g^T$. Written as a matrix, this gives:

$$\Lambda = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^n \\ \alpha & 1 + \alpha^2 & \alpha(1 + \alpha^2) & \dots & \alpha^{n-1}(1 + \alpha^2) \\ \alpha^2 & \alpha(1 + \alpha^2) & 1 + \alpha^2 + \alpha^4 & & \\ \vdots & \vdots & & \ddots & \\ \alpha^n & \alpha^{n-1}(1 + \alpha^2) & & & 1 + \dots + \alpha^{2n} \end{bmatrix}. \quad (17)$$

The n th term along the diagonal is

$$\Lambda(n, n) = \sum_{j=0}^n \alpha^j = \frac{1 - \alpha^{2n}}{1 - \alpha^2}$$

and off-diagonal terms $\Lambda(i, j) = \Lambda(i, i)\alpha^{j-i}$, for $i < j$. We can also express Λ analytically in continuous time. Diagonal terms are given by:

$$\Lambda(t, t) = \int_0^t e^{-2gs} ds = \frac{1}{2g}(1 - e^{-2gt})$$

and off-diagonal terms by $\Lambda(t, t') = e^{-g(t'-t)}\Lambda(t, t)$, for $t < t'$.

References

- [1] J. W. Pillow, L. Paninski, and E. P. Simoncelli. Maximum likelihood estimation of a stochastic integrate-and-fire neural model. In L. Saul S. Thrun and B. Scholkopf, editors, *Advances in Neural Information Processing Systems*, volume 16, 2004.
- [2] J. W. Pillow. *Neural coding and the statistical modeling of neuronal responses*. PhD thesis, NYU, 2005.
- [3] L. Paninski, J. W. Pillow, and E. P. Simoncelli. Maximum likelihood estimation of a stochastic integrate-and-fire neural model. *Neural Computation*, 16:2533–2561, 2004.
- [4] J. W. Pillow, L. Paninski, V. J. Uzzell, E. P. Simoncelli, and E. J. Chichilnisky. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *The Journal of Neuroscience*, 25:11003–11013, 2005.
- [5] W. Gerstner and W. Kistler. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, 2002.
- [6] J. W. Pillow and E. P. Simoncelli. Biases in white noise analysis due to non-Poisson spike generation. *Neurocomputing*, 52:109–115, 2003.
- [7] S. Karlin and H. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 1981.
- [8] B. Knight, A. Omurtag, and L. Sirovich. The approach of a neuron population firing rate to a new equilibrium: an exact theoretical result. *Neural Computation*, 12:1045–1055, 2000.
- [9] L. Paninski, B. Lau, and A. Reyes. Noise-driven adaptation: in vitro and mathematical analysis. *Neurocomputing*, 52:877–883, 2003.
- [10] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical recipes in C*. Cambridge University Press, 1992.